

Секция «Иностранные языки и регионоведение»

Фразеографические проекты в корпусной лингвистике

Кротова Елена Борисовна

Аспирант

Московский государственный университет имени М.В. Ломоносова, Факультет
иностранных языков и регионоведения, Москва, Россия

E-mail: elena_krotova@inbox.ru

Корпусы текстов, снабжённые машиночитаемой разметкой и специализированной поисковой системой, появились не так давно. Первым из них считается Брауновский корпус, созданный в 1960-е годы. В 1969 г. вышел словарь “The American Heritage Dictionary of the English Language”, использовавший данные корпуса, в частности, для определения частотности слов. В 1987 г. появился словарь Collins COBUILD English Language Dictionary, базировавшийся на данных корпуса “The Bank of English”. Дальнейшее развитие компьютерных технологий способствовало росту числа лексикографических исследований, в большей либо меньшей степени использующих данные машиночитаемых корпусов.

Исследователи выделяют два подхода к использованию корпусов в лексикографических исследованиях: базирующийся на корпусах (corpus-based) и направляемый корпусом (corpus-driven) [1; 3; 4]. В рамках первого подхода корпусные данные используются для подтверждения уже существующей гипотезы, в рамках второго подхода данные о языковых структурах предоставляет проводимый машиной анализ, результаты которого затем интерпретируются лингвистом. В моем докладе я проиллюстрирую на примере двух лексикографических проектов, как данные подходы могут реализовываться на практике.

Первым лексикографическим проектом, о котором пойдёт речь, будет проект "EPHRAS - Ein mehrsprachiges phraseologisches Lernmaterial" целью которого является развитие и составление многоязычного фразеологического учебного материала для немецкого, словенского, словацкого и венгерского языков [2; 5]. Все словарные статьи построены по единой модели и содержат информацию о формально-структурных, грамматических, семантических и стилистическо-прагматических свойствах фразеологизмов. Корпусные данные использовались, в частности, при определении формы леммы и возможных вариантов фразеологизмов, их лексической сочетаемости, валентностей; каждая статья снабжена контекстами, иллюстрирующими приводимую информацию. В докладе будет показано, что проект можно рассматривать как образец фразеографического исследования, проводимого в рамках первого подхода.

Получить представление о значимости корпусного анализа для лексикографии можно, сравнив словарную статью для идиомы *nicht alle Tassen im Schrank / Spind haben* из словаря Duden. Redewendungen“ со статьёй из вышеуказанного проекта. В словарной статье лемма снабжена пометой «разг.», толкованием и двумя контекстами. В словарной статье проекта, помимо этого, более подробно описана семантика идиомы, указаны её валентность, синтаксическая функция, лексическая сочетаемость, дан комментарий по прагматическим свойствам и имеются шесть контекстов употребления идиомы. То есть информация является более полной и обеспечивает не только понимание идиомы, но

и способствует её безошибочному употреблению пользователем, изучающим любой из языков проекта.

Примером реализации на практике второго подхода может служить проект "Usuelle Wortverbindungen над которым ведётся работа в Институте немецкого языка в Мангейме [1]. Принципиальным его отличием от проекта EPHRAS является то, что работа лингвиста направляется данными корпусного анализа, и происходит это следующим образом: машина анализирует, какие слова со среднестатистической точки зрения чаще, чем другие, встречаются вместе, после чего выявленные языковые структуры интерпретируются лингвистом. Предполагается, что таким образом можно получить новые сведения о языке и его фразеологической системе, в том числе о «новых» и/или ещё не зафиксированных в словарях узуализированных сочетаниях (usuelle Wortverbindungen), их семантике, возможных модификациях, типичных контекстах и возможных связях с другими словосочетаниями. Примеры словарных статей, создаваемых в рамках проекта при помощи анализа сочетаемости и иных данных, получаемых из корпусов, опубликованы на сайте экспериментального проекта "Wortverbindungen online"[6]. Характерным является то, что представлены все классы фразеологизмов, обладающие различной степенью идиоматичности, так как основанием для их выбора являлась не принадлежность к определённому классу фразеологизмов, а частотность их совместной встречаемости (Kookkurrenz) в корпусах.

В моем докладе будет рассмотрен вопрос о том, как машиночитаемые корпуса могут быть применены во фразеографической практике, и на примере вышеописанных проектов проиллюстрировано их значение для лексикографии.

Литература

1. Brunner A., Steyer K. Phraseologische und phraseographische Aspekte korpusgesteuerter Empirie. In: Jesenšek, Vida/Fabčič, Melanija (Hrsg.): Phraseologie kontrastiv und didaktisch. Neue Ansätze in der Fremdsprachenvermittlung, Maribor: Slavistično, društvo: Filozofska fakulteta 2007, S. 181-194.
2. Jesenšek V. Phraseologische Wörterbücher auf dem Weg zu Phraseologiedatenbanken. In: Mellado Blanco, Carmen (Hrsg.): Theorie und Praxis der idiomatischen Wörterbücher. (Lexicographica. Series maior/ Supplementbände zum Internationalen Jahrbuch für Lexikographie 135), Tübingen, 2009, S. 65-82.
3. Lemnitzer L., Zinsmeister H. Korpuslinguistik. Eine Einführung. Narr Studienbücher, 2006.
4. Sinclair, J. Corpus Concordance Collocation. Oxford University Press. 1991
5. EPHRAS - Ein mehrsprachiges phraseologisches Lernmaterial auf CD-ROM. URL: <http://www.ephras.org/index.php>
6. Wortverbindungen online. URL: <http://hypermedia.ids-mannheim.de/elexiko/wv-online/>
7. Duden (2008): Redewendungen. Wörterbuch der deutschen Idiomatik. Mannheim.